

AD-A255 991



12

AN ITERATIVE MONTE CARLO METHOD FOR
NONCONJUGATE BAYESIAN ANALYSIS

by

Bradley P. Carlin

Alan E. Gelfand

DTIC

ELECTE

SEP 21 1992

C

TECHNICAL REPORT No. 457

SEPTEMBER 7, 1992

Prepared Under Contract

N00014-92-J-1254 ((NR-042-267))

FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

Approved for public release; distribution unlimited

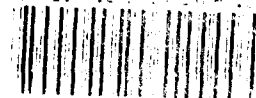
DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065



92 9 18 044



9225499

AN ITERATIVE MONTE CARLO METHOD FOR
NONCONJUGATE BAYESIAN ANALYSIS

by
Bradley P. Carlin
Alan E. Gelfand

TECHNICAL REPORT No. 457
SEPTEMBER 7, 1992

Prepared Under Contract
N00014-92-J-1254 ((NR-042-267))
FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

INTC QUALITY INSPECTED 3

| | |
|--------------------|-------------------------------------|
| Accession For | |
| NTIS | <input checked="" type="checkbox"/> |
| DTIC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| Distribution/ | |
| Availability Codes | |
| Avail and/or | |
| Dist | Special |
| A-1 | |

An Iterative Monte Carlo Method for
Nonconjugate Bayesian Analysis

Bradley P. Carlin
Carnegie Mellon University

Alan E. Gelfand
University of Connecticut

ABSTRACT

The Gibbs sampler has been proposed as a general method for Bayesian calculation in Gelfand and Smith (1990). However experience to date is almost exclusively in applications assuming conjugacy where implementation is reasonably straightforward. This paper describes a tailored rejection method approach for implementation of the Gibbs sampler when nonconjugate structure is present. Several challenging applications are presented for illustration.

Keywords: Bayesian inference, Gibbs sampler, hierarchical models, logistic regression, nonlinear models, rejection method

AMS 1980 Subject Classification: 62 F 15, 65 C 05, 10

1. Introduction

In earlier work (Gelfand and Smith, 1990; Gelfand et. al., 1989) a sampling based approach using the Gibbs sampler (Geman and Geman, 1984) was offered as a means for implementing Bayesian data analysis. This approach is very broadly applicable but experience to date resides almost exclusively in applications assuming conjugacy. Two notable exceptions are Zeger and Karim (1989) and Racine-Poon et. al. (1990). By way of clarification, in the context of a hierarchical Bayes model, conjugacy is taken to mean that for any parameter in the model specification (likelihood * prior), integration of this model with respect to this parameter may be achieved explicitly. This pragmatic definition differs a bit from formal versions in e.g. Diaconis and Ylvisaker (1979) or in Morris (1983).

Conjugacy allows us to simplify the implementation of the Gibbs sampler enabling almost routine fully Bayesian analysis of many standard problems. However, more challenging modeling situations will not allow conjugacy as the following examples suggest:

- (i) reference priors (Bernardo, 1979; Berger and Bernardo, 1989) and other versions of "noninformative" priors (Berger, 1985) will not be conjugate with the likelihood.
- (ii) nonlinear models, including generalized linear models, will have likelihoods as functions of the model parameters which typically do not admit a conjugate form.
- (iii) for hierarchical models, according to Berger (1985, p. 232) "the choice of a form for the second or higher stage prior seems to have relatively little effect". However, this is usually not the case at the first stage of prior specification where the form of the prior, e.g., body and tails will have substantial effect on the inference. To assess model robustness requires fully Bayesian analysis when the first stage prior is nonconjugate.

To carry out the Gibbs sampler in the presence of nonconjugacy for at least some of the model parameters requires sampling from nonstandardized densities as discussed below. A means for accomplishing such sampling is the rejection method (Devroye, 1986; Ripley, 1986). The rejection method is formally defined in Section 3. The purpose of this paper is to describe a tailored general rejection method approach for implementation of the Gibbs sampler when some nonconjugate structure is present.

To clarify how nonstandardized densities arise we note that the Gibbs sampler requires independent draws from the complete conditional distributions of the model parameters (see Sec. 2). For any parameter in any hierarchical model, its complete conditional distribution is the conditional distribution of the parameter given the data and all other model parameters. But, it is then clear that for each model parameter its complete conditional density is proportional to likelihood * prior. Often the hierarchical

structure along with, for example, exchangeability assumptions greatly simplify these distributions.

In the next section we briefly review the Gibbs sampler. Since it is a replicated iterative Markovian updating scheme, the conditional levels for each complete conditional distribution which needs to be sampled will change with each iteration and replication. Standard use of the rejection method requires that a distinct envelope function be developed for each such sampling. Unfortunately this envelope is used to generate but one observation. As an alternative, in Section 3 we first note that a good multidimensional envelope density will provide good complete conditional envelope densities. We then show how to create such a multidimensional envelope density which also possesses complete conditional distributions that are easy to sample.

In Section 4 we illustrate with three demanding modeling applications. Finally in Section 5 we summarize noting, in addition, when our proposed method is likely to work well and when not.

2. The Gibbs Sampler

For convenience, in this section densities will be denoted generically by square brackets, so that joint, conditional and marginal forms for random variables U, V , appear as $[U, V]$, $[U|V]$ and $[V]$, respectively. The usual marginalization by integration is denoted by forms such as $[U] = \int [U|V] \cdot [V]$. For a collection of random variables $[U_1, U_2, \dots, U_k]$ the complete conditional densities can thus be denoted by $[U_s|U_r, r \neq s]$, $s = 1, 2, \dots, k$, and the marginal densities by $[U_s]$, $s = 1, 2, \dots, k$.

Given the ability to draw random variate samples of U_s from $[U_s|U_r, r \neq s]$ for specified values of $\{U_r, r \neq s\}$, $s = 1, 2, \dots, k$, the Gibbs sampler provides an iterative scheme which enables us to make sample-based estimates, $[\hat{U}_s]$, of the marginal densities, $[U_s]$, $s = 1, 2, \dots, k$. The scheme proceeds as follows: given an arbitrary starting set of values $U_1^{(0)}, \dots, U_k^{(0)}$, we draw $U_1^{(1)}$ from $[U_1|U_2^{(0)}, \dots, U_k^{(0)}]$, then $U_2^{(1)}$ from $[U_2|U_1^{(1)}, U_3^{(0)}, \dots, U_k^{(0)}]$, and so on up to $U_k^{(1)}$ from $[U_k|U_1^{(1)}, \dots, U_{k-1}^{(1)}]$ to complete one iteration of the scheme. After t such iterations we would arrive at joint a sample $(U_1^{(t)}, \dots, U_k^{(t)})$. Geman and Geman (1984) show, under mild conditions, that $(U_1^{(t)}, \dots, U_k^{(t)}) \xrightarrow{d} (U_1, \dots, U_k)$ - $[U_1, U_2, \dots, U_k]$ as $t \rightarrow \infty$. Hence for t large enough, $U_s^{(t)}$, for example, can be regarded as

a sample variate from $[U_s]$. Parallel replication of this process m times yields m iid k -tuples $(U_{1j}^{(t)}, \dots, U_{kj}^{(t)})$, $j = 1, 2, \dots, m$. Note that sample size at, say, the w -th iteration may be increased from m to any specified size by randomly reusing the $U_{sj}^{(w-1)}$ with replacement.

It is shown in Gelfand and Smith (1990) that a density estimate of the form

$$[\hat{U}_s] = \frac{1}{m} \sum_{j=1}^m [U_s | U_{rj}^{(t)}, r \neq s] \quad (1)$$

is better than a kernel density estimate for $[U_s]$. This is not surprising since (1) takes advantage of the known structure in the model whereas the kernel density estimate does not. The form (1) is a discrete mixture distribution and is essentially a Monte Carlo integration to accomplish the desired marginalization. Extension to expectations, $E[h(U_s)]$, and more generally to densities and expectations for functions $W(U_1, U_2, \dots, U_k)$ is straightforward (see Gelfand and Smith, 1989 for details).

In the Bayesian context the U_i are the unknown parameters (or possibly unobserved data) in the model, and W would be any function of the parameters (or unobserved data) which is of interest. All distributions are viewed as conditional on the observed data, whence the marginal densities, $[U_s]$, become the desired marginal posterior distributions of the parameters (or unobserved data). Moreover, the joint density $[U_1, \dots, U_k]$ becomes the joint density of all the model parameters/unknowns given the observed data. This density, only known modulo normalizing constant, will be denoted by $f(U_1, \dots, U_k)$ where f is, in fact, likelihood \times prior. Similarly all complete conditional distributions will again be proportional to f and, in the absence of conjugacy, will not lead to familiar standard forms such as normals and gammas. As a result, though the forms of the densities required for use of the Gibbs sampler are known, sampling will require random generation from nonstandardized densities. In Section 3 we suggest an approach to accomplish this using a tailored version of the rejection method.

As noted above we prefer to use a density estimate of the form (1). In fact using this form allows m to be much smaller (say $m = 100$) than needed for kernel density estimates (say $m = 5000$). However, calculation of (1) will require, at the last iteration, m normalizations of f which in turn requires m one-dimensional numerical integrations. Simple trapezoidal or Simpson's rule integration to do this is quite fast still yielding

substantial overall savings in run time compared with kernel density estimation.

Finally, we note that complete implementation of the Gibbs sampler requires that a determination of t be made and that, across iterations, choice(s) of m be specified. In a challenging application some experimentation with different settings for t and m will likely be necessary. We do not view this as a deterrent since random generation is generally inexpensive and since in many cases there may be no feasible alternative. In the examples of Section 4, convergence was evaluated in a univariate manner by plotting marginal posterior density estimates of the form (1) five iterations apart in order to judge stability. Typically, a somewhat small m is used until convergence is concluded, at which point, for a final iteration, m is increased by an order of magnitude to develop the presented density estimate. We make no claims for the optimality of this procedure. Assessment of convergence is a complex issue which is currently very much in the empirical domain.

3. A Tailored Rejection Method

In this section we develop a specialized version of the rejection method which is well-suited to the sampling needs of the Gibbs sampler. First we review the basic rejection algorithm method.

3.1 The Rejection Algorithm

The rejection algorithm for a nonstandardized integrable density $f(\underline{U})$, $\underline{U} = (U_1, \dots, U_k)$ proceeds as follows.

- (i) Identify a density $g(\underline{U})$ which may be readily sampled and such that there exists M for which $f(\underline{U})/g(\underline{U}) \leq M$ for all \underline{U} .
- (ii) Generate \underline{U}^* from $g(\underline{U})$.
- (iii) Generate V from a $U(0,1)$ distribution.
- (iv) Accept \underline{U}^* if $V \leq f(\underline{U}^*)/Mg(\underline{U}^*)$, otherwise return to (ii).

It may be shown (see Devroye, 1986 or Ripley, 1986) that the distribution of \underline{U}^* is $f(\underline{U})/\int f(\underline{U})$ and also that the acceptance probability associated with this algorithm is $M^{-1} \int f(\underline{U})$. Hence the smaller we can make M , that is, the more g resembles f , the more

efficient the sampling.

3.2 Split-normal and Split-t Envelope Functions

Implementation of the Gibbs sampler requires sampling from f viewed as a function of, say, U_1 for fixed $\underline{U}_{-1} \equiv (U_2, \dots, U_k)$. However the value of \underline{U}_{-1} changes with each iteration and each replication. Customary use of the rejection method then requires that a distinct envelope function $g_{\underline{U}_{-1}}(U_1)$ be developed for each \underline{U}_{-1} . Moreover, typically each such $g_{\underline{U}_{-1}}(U_1)$ is used to generate but one observation.

As an alternative, now viewing f as a k -dimensional function we propose, before doing any sampling, to create a single k -dimensional density function $g(\underline{U})$ which is a good envelope for f and is such that for each U_i , g has complete conditional distributions which are easy to sample. Formalizing notation and still taking $i = 1$ we write $g(U_1, \dots, U_k) = g_1(U_1 | \underline{U}_{-1}) \cdot g_2(\underline{U}_{-1})$. Note that $g_1(U_1 | \underline{U}_{-1})$ serves as an envelope for the complete conditional distribution for U_1 arising from f . That is, if M is such that $f(\underline{U})/g(\underline{U}) \leq M$ for all \underline{U} then, as a function of U_1 for fixed \underline{U}_{-1} , $f/g_1 \leq M' \equiv M g_2(\underline{U}_{-1})$. In practice g_1 , g_2 and M' are not calculated; acceptance of U_1^* is determined by the equivalent test (iv) above evaluating f and g at $(U_1^*, \underline{U}_{-1})$.

How might a suitable $g(\underline{U})$ be developed? Writing $f(\underline{U}) = \text{likelihood}(\underline{U}) \cdot \text{prior}(\underline{U})$, if $\hat{\underline{U}}$ is the maximum likelihood estimate of \underline{U} we may take $g(\underline{U}) = \text{prior}(\underline{U})$ with $M = \text{likelihood}(\hat{\underline{U}})$ to implement the rejection method. This choice of g has at least two drawbacks. First, since it only matches the prior, it need not be a good envelope for f so that very inefficient sampling may result. Second, it requires $\text{prior}(\underline{U})$ to be proper (since we sample from g in the rejection method) where as, in fact, we only need f proper. Hence, while this choice of g may be viewed as a possible backup we seek a proper g which more closely resembles f .

In the context of noniterative Monte Carlo sampling with respect to a nonstandardized density, Geweke (1989) proposes the use of an importance sampling density which is a multivariate split-normal or split-t distribution. Such a density, g , is designed to make the variability of the ratio f/g over the space of \underline{U} small under g

which in turn makes the variance of the Monte Carlo integration small. Note that such a g is desirable for our purposes since the less variable f/g is, the smaller M will be whence the greater the acceptance probability and the more efficient our sampling.

Recall that in the Bayesian framework, modulo normalization, f is viewed as the joint posterior density function of all the parameters (and perhaps any missing data) given the observed data. With an increasing amount of data, under usual regularity conditions f is approximately a multivariate normal density up to a proportionality constant (see e.g. Berger 1985, p. 224). A convenient choice to approximate the mean of this normal distribution is \hat{U} , the mode of f . With regard to an approximate covariance matrix, the preferred choice from an asymptotic viewpoint is the negative of the inverse Hessian evaluated at \hat{U} . The Hessian matrix H is defined by $H_{ij} = \partial^2 \log f / \partial U_i \partial U_j$. In two stage models we might use the log likelihood rather than $\log f$ which amounts to replacing H by the information matrix I .

Often both the H matrix and the I matrix are difficult to obtain since they require the existence of and the evaluation of second derivatives. A differencing algorithm (such as in Kass, 1987) can be used to provide reliable derivative-free estimates for the matrices H or I thus avoiding formal differentiation. Since our objective is only to approximate the covariance matrix, Σ , associated with f we need not use these asymptotic forms but may instead adopt alternative choices for $\hat{\Sigma}$. One simple approach which avoids the differentiation problem is to approximate the surface $\log f(\underline{U})$ by a quadratic function, e.g. $(\underline{U} - \hat{\underline{U}})^T V (\underline{U} - \hat{\underline{U}})$ whence $\hat{\Sigma} = -V^{-1}$. This approximation can be straightforwardly developed by usual least squares methods fitting the quadratic to a large set of $\log f$ values obtained by evaluating f at many points on a k -dimensional grid. It may in fact prove easiest to first transform \underline{U} to the k -dimensional unit square, obtain the covariance matrix estimate and then transform this estimate back to the original scale by the delta method. When there are strong correlations amongst the U 's or when $\log f$ is fairly flat H , I and V may be nearly singular making inversion awkward. This problem can be alleviated appropriate reparametrization, i.e., transformation of \underline{U} .

We note another approach which avoids both differentiation and inversion problems but at the expense of computational effort that will become infeasible with increasing dimensionality. We can obtain a piecewise uniform approximation to f and then obtain the covariance matrix associated with this approximation. For simplicity of illustration assume f is bivariate. Again, it may be easiest to transform (U_1, U_2) to the unit square

with resulting density proportional to h . Partition the unit square into a grid of r^2 cells and evaluate h at say the midpoint of each cell in the grid. Denoting these values by h_{ij} , $i, j = 1, \dots, r$, replace the density h by the constant h_{ij} for points in the $(i,j)^{th}$ cell to obtain the piecewise uniform approximation to h . Normalization of this approximate density and calculation of its moments is straightforward. Thus we may approximate the covariance matrix associated with h and, again using the delta method, that associated with f .

The above discussion suggests taking $g(\underline{U})$ to be $N(\hat{\underline{U}}, \hat{\underline{\Sigma}})$ for some convenient $\hat{\underline{\Sigma}}$. However for more interesting situations involving small to moderate amounts of data, although \hat{f} will typically still be unimodal, it will likely be somewhat asymmetric and our choice of $\hat{\underline{\Sigma}}$ will likely be a weak covariance approximation. Geweke (1989) suggests that an appropriate split-normal or split-t distribution be used in place of $N(\hat{\underline{U}}, \hat{\underline{\Sigma}})$. We now develop the details of this approximation for our situation including the required complete conditional distributions.

A standard univariate split-normal distribution denoted by $SN(0, q, r)$ is defined by the density

$$\begin{cases} \frac{1}{\sqrt{2\pi} q} e^{-z^2/2q^2}, & z > 0 \\ \frac{1}{\sqrt{2\pi} r} e^{-z^2/2r^2}, & z < 0 \end{cases}$$

To generate $Z \sim SN(0, q, r)$ we draw $\epsilon \sim \eta(0, 1)$ and take $Z = q\epsilon$ if $\epsilon > 0$, $Z = r\epsilon$ if $\epsilon < 0$. Let $\underline{Z}^t = (Z_1, \dots, Z_k)$ be a random vector such that Z_i are independent with $Z_i \sim SN(0, q_i, r_i)$. A general multivariate split normal arises by affine transformation of \underline{Z} . In particular, it is proposed to take g to be the distribution of $\underline{U} = \hat{\underline{U}} + \hat{\underline{\Sigma}}^{1/2} \underline{Z}$ for q_i, r_i given below.

Choices for q_i and r_i are intended to make g a better envelope than $N(\hat{\underline{U}}, \hat{\underline{\Sigma}})$. Geweke proposes that $q_i = \sup_{\Delta > 0} v_i(\Delta)$, $r_i = \sup_{\Delta < 0} v_i(\Delta)$ where

$$v_i(\Delta) = \frac{|\Delta|}{\sqrt{2(\log f(\hat{\underline{U}}) - \log f(\hat{\underline{U}} + \Delta \hat{\underline{\Sigma}}^{1/2} \underline{\lambda}^{(i)}))}} \quad (2)$$

and $\lambda^{(i)}$ is a unit vector in the i^{th} coordinate direction. Geweke notes that the Δ 's yielding these maxima correspond to the positive and negative values respectively along the i^{th} coordinate axis which maximize the ratio of the rate of decline of f , $f(\hat{U} + \Delta \hat{\Sigma}^{1/2} \lambda^{(i)})/f(\hat{U})$, to the rate of decline of g , $g(\hat{U} + \Delta \hat{\Sigma}^{1/2} \lambda^{(i)})/g(\hat{U})$. Choice of q_i, r_i in this manner gives f/g the same value at Δ such that $q_i = \sup_{\Delta > 0} (V_i(\Delta))$ ($r_i = \sup_{\Delta < 0} (V_i(\Delta))$) as at $\Delta = 0^+(0^-)$. Such matching aids in making f/g "more constant" in each coordinate direction. Exact calculation of q_i, r_i is an analytical problem generally without explicit solution. Practically, these values are obtained only approximately by evaluating $v_i(\Delta)$ over the set $\{\Delta = j/2, j = \pm 1, \pm 2, \dots, \pm 12\}$.

The tail behavior associated with the form of f might suggest that a better choice for g would be a multivariate split- t distribution. A standard univariate split- t with ν d.f., $ST(\nu; 0, q, r)$ arises as the distribution of $t = Z/\sqrt{V/\nu}$ where $Z \sim SN(0, q, r)$ independent of V , a χ^2 random variable with ν d.f. To generate $t \sim ST(\nu; 0, q, r)$ we draw $\xi \sim t_\nu$ and take $t = q\xi$ if $\xi > 0$, $t = r\xi$ if $\xi < 0$. More generally let $\underline{t} = (t_1, \dots, t_k)$ be a random vector where $t_i = Z_i/\sqrt{V/\nu}$ with Z_i independent, $Z_i \sim SN(0, q_i, r_i)$ independent of $V \sim \chi^2_\nu$. A general multivariate split- t arises by affine transformation of \underline{t} . In particular it is proposed to take g to be the distribution of $\underline{U} = \hat{\underline{U}} + \hat{\Sigma}^{1/2} \underline{t}$ with q_i and r_i calculated, replacing $v_i(\Delta)$ in (2) with

$$v_i(\Delta) = \frac{|\Delta|}{\sqrt{\nu((f(\hat{\underline{U}})/f(\hat{\underline{U}} + \Delta \hat{\Sigma}^{1/2} \lambda^{(i)}))^2/(k+\nu)-1)}} \quad (3)$$

Remarks below (2) are applicable here.

We comment that it seems preferable to transform (reparametrize) each U_i to have R^1 as support before embarking on the creation of g to "match" f .

Returning to the multivariate split-normal it is perhaps easiest to think of the transformation from \underline{Z} to \underline{U} as arising from 2^k one-to-one transformations determined by the vector $\text{sgn} \underline{Z} = (\text{sgn } Z_1, \text{sgn } Z_2, \dots, \text{sgn } Z_k)$. Index these transformations by $j = 1, 2, \dots, 2^k$ with associated partitions of R^k denoted by A_j . On A_j there will be an associated set of q 's and r 's. In fact let $d_{ji} = q_i^2$ if $\text{sgn } Z_i = 1$ on A_j , r_i^2 if $\text{sgn } Z_i = -1$

on A_j and let D_j be a diagonal matrix with diagonal entries d_{jj} . Then on A_j , $\underline{Z} \sim N(0, D_j)$ and thus the density for \underline{Z} is, in obvious notation

$$h(\underline{Z}) = \sum_{j=1}^k N(0, D_j)(\underline{Z}) \cdot 1_{A_j}(\underline{Z}) \quad (4)$$

If B_j is the image of A_j under the transformation $\underline{U} = \hat{\underline{U}} + \hat{\Sigma}^{1/2} \underline{Z}$ then the density for \underline{U} is

$$g(\underline{U}) = \sum_{j=1}^k N(\hat{\underline{U}}, \hat{\Sigma}^{1/2} D_j (\hat{\Sigma}^{1/2})^t)(\underline{U}) \cdot 1_{B_j}(\underline{U}) \quad (5)$$

The Gibbs sampler requires sampling from the complete conditional distributions associated with \hat{f} . By earlier remarks, this requires sampling from the complete conditional distributions associated with g . But, for example, what is $g(U_1 | \underline{U}_{-1})$ for the density (5)? We now show that this distribution is a univariate split normal which can be easily sampled. Choose for $\hat{\Sigma}^{1/2}$ the upper triangular (Cholesky) decomposition of $\hat{\Sigma}$ which we denote by

$$T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$

Note that, by using T , \underline{U}_{-1} uniquely determines $\underline{Z}_{-1} = T_{22}^{-1}(\underline{U}_{-1} - \hat{\underline{U}}_{-1})$. Furthermore $U_1 = W_1(\underline{U}_{-1}) + T_{11}Z_1$ where $W_1(\underline{U}_{-1}) = \hat{U}_1 + T_{12}T_{22}^{-1}(\underline{U}_{-1} - \hat{\underline{U}}_{-1})$. Hence $U_1 | \underline{U}_{-1} = U_1 | \underline{Z}_{-1} = (W_1(\underline{U}_{-1}) + T_{11}Z_1) | \underline{Z}_{-1} = W_1(\underline{U}_{-1}) + T_{11}Z_1$ i.e. $U_1 | \underline{U}_{-1}$ has a univariate split normal distribution. Moreover $U_1 | \underline{U}_{-1}$ is easily sampled by drawing $Z_1 \sim SN(0, q_1, r_1)$ and making the linear transformation $W_1(\underline{U}_{-1}) + T_{11}Z_1$. We remark that T and $\hat{\underline{U}}$ do not change from iteration to iteration, from replication to replication. Given \underline{U}_{-1} we need only calculate $W(\underline{U}_{-1})$ which just involves linear operations on \underline{U}_{-1} .

In the multivariate split-t case $\underline{t} = (t_1, \dots, t_k)$ arises from $t_i = Z_i / \sqrt{V/\nu}$ with Z_i

independent, $Z_i \sim \text{SN}(0, q_i, r_i)$ independent of $V \sim \chi^2_\nu$. Then, analogous to (4), with obvious notation

$$h(\underline{t}) = \sum_{j=1}^{2^k} t_{\nu(0, D_j)}(\underline{t}) \cdot 1_{A_j}(\underline{t}) \quad (6)$$

and for $\underline{U} = \hat{\underline{U}} + \hat{\Sigma}^{1/2} \underline{t}$,

$$g(\underline{U}) = \sum_{j=1}^{2^k} t_{\nu(\hat{\underline{U}}, \hat{\Sigma}^{1/2} D_j (\hat{\Sigma}^{1/2})^t)}(\underline{U}) \cdot 1_{B_j}(\underline{U}) \quad (7)$$

Again we require associated complete conditional distributions. Careful evaluation shows that $U_1 | \underline{U}_{-1}$ now has a univariate split-t distribution. More precisely $U_1 | \underline{U}_{-1} \sim W_1(\underline{U}_{-1}) + T_{11} V_1(\underline{U}_{-1}) t_1$ where $t_1 \sim \text{ST}(\nu+k-1, 0, q_1, r_1)$ and $V_1(\underline{U}_{-1}) = ((\nu+k-1)^{-1} (\nu + \sum_{i=1}^k Z_i^2 / e_i))^{1/2}$ with the Z_i being components of \underline{Z}_{-1} defined above and $e_i = q_i^2$ or r_i^2 according to whether $Z_i > 0$ or < 0 . Given \underline{U}_{-1} we need to calculate $V_1(\underline{U}_{-1})$ in addition to $W_1(\underline{U}_{-1})$.

We commented earlier, that in implementing the rejection method we would test a U_1^* generated from $U_1 | \underline{U}_{-1}$ by using (iv) in Section 3.1. Computation is simplified by noting that $g(\underline{U}) = \prod_{i=1}^k T_{ii}^{-1} \cdot h(T^{-1}(\underline{U} - \hat{\underline{U}}))$ with h as in (4) or in (6) accordingly.

However it still remains to choose M . It seems natural to look at the ratio f/g at the mode $\hat{\underline{U}}$ but, as yet, g is undefined at $\hat{\underline{U}}$ i.e. h is undefined at $\underline{0}$. Let $h(\underline{0}) =$

$a \prod_{i=1}^k \min(q_i^{-1}, r_i^{-1})$ where in the split-normal case $a = (2\pi)^{-k/2}$ while in the split-t case

$a = \Gamma(\frac{\nu+k}{2}) / (\Gamma(\nu/2)(\pi\nu)^{k/2})$. Then $g(\hat{\underline{U}}) = \prod_{i=1}^k T_{ii}^{-1} \cdot h(\underline{0})$. Define $M(\hat{\underline{U}}) = f(\hat{\underline{U}})/g(\hat{\underline{U}})$.

For both the split-normal and split-t cases as a result of the way h was chosen along with its definition at $\underline{0}$ $M(\hat{\underline{U}})$ will bound f/g in a neighborhood of $\hat{\underline{U}}$. In practice, choosing $M = bM(\hat{\underline{U}})$ with $1.2 \leq b \leq 5$ has provided an overall bound for f/g . In our experience choice of d.f. (ν) to accommodate the tail behavior of f is more critical.

If, during the course of sampling, a \underline{U}_0 arises such that $f(\underline{U}_0)/g(\underline{U}_0)$ violates our bound we do a local search in a neighborhood of \underline{U}_0 and revise M accordingly. Before M is revised the magnitude of $f(\underline{U}_0)/Mg(\underline{U}_0)$ provides a rough idea of the severity of the violation. Of course if a violation occurs then some of the previously generated variates might not have been retained with this revised M and more importantly we were not sampling from the desired complete conditional distribution. Before exploring this point further suppose that the change in M is small (as is typically the case in our experience) so that most of the previously generated variates would still be retained. Then we would expect the joint distribution of \underline{U} at the current iteration of the sampler to be closer to the converged joint distribution than when we started. Thus we would expect no advantage to starting the sampler anew as opposed to continuing from the current iteration.

Continuing with these ideas, suppose for a given M we define $S_M = \{\underline{U} : f(\underline{U})/g(\underline{U}) > M\}$. Following the argument which justifies the rejection method we may show that the distribution of \underline{U} is actually

$$\frac{f(\underline{U})}{\int_{S_M^c} f(\underline{U})d\underline{U} + MP_g(S_M)}, \quad \underline{U} \in S_M^c$$

$$\frac{Mg(\underline{U})}{\int_{S_M^c} f(\underline{U})d\underline{U} + MP_g(S_M)}, \quad \underline{U} \in S_M \quad (3)$$

Unfortunately $\int_{S_M} f(\underline{U})d(\underline{U}) > MP_g(S_M)$ so that even if $P_g(S_M)$ is very small we cannot be sure that (8) is close to $f(\underline{U})/\int f(\underline{U})d\underline{U}$. Hence complete conditional distributions arising from (8) need not be close to complete conditional distributions arising from $f(\underline{U})$. More optimistically, if for example given \underline{U}_{-1} the set of \underline{U}_1 such that $f(\underline{U})/g(\underline{U}) > M$ is a null set then we are, in fact, sampling from the complete conditional distribution of $\underline{U}_1 | \underline{U}_{-1}$ arising from f .

We conclude this section with an important remark. When k is large,

development of $g(\underline{U})$ will be made difficult because of complications in obtaining $\hat{\underline{U}}, \hat{\Sigma}$ and T . However, in most applications $f(\underline{U})$ is a product of functions. Hence, if we need to sample from f viewed as a nonstandardized density for $U_1 | \underline{U}_{-1}$ we need only consider the terms in this product involving U_1 and only the variables say $U_2, \dots, U_{k'}$, appearing in these terms. That is, we factor $f(\underline{U})$ as $f(U_1, \dots, U_k) = f_1(U_1, U_2, \dots, U_{k'}) \cdot f_2(\underline{U}_{-1})$ so that g need only be a k' dimensional envelope function. Typically, k' is much smaller than k as, for instance, in exchangeable models.

4 Examples

In this section we apply the tailored rejection method to three nonconjugate modeling scenarios. Each has been chosen to illustrate one or more of points (i)–(iii) of Section 1.

4.1 Asymptotic Regression Model

Consider a model having mean structure

$$E(Y_i) = \alpha - \beta \gamma^{X_i}, \quad \alpha, \beta > 0, \quad 0 < \gamma < 1 \quad (9)$$

This equation describes a growth curve which has no inflection point and approaches an asymptote as X_i tends to infinity. Models of this type find agricultural, biological, and engineering application. To complete the specification of the model, we assume independent $Y_i \sim N(E(Y_i), \sigma^2)$, $i=1, \dots, n$, and adopt the vague prior $\pi(\alpha, \beta, \gamma, \sigma) \propto (\alpha\sigma)^{-1}$ considered by Hills (1989). While the prior we have adopted is not a reference prior in the sense of Bernardo (1979), it is a vague prior in the spirit of (i) in Section 1. In any event, the nonlinear structure in (9) precludes conjugate priors as noted in (ii) in Section 1.

In order to implement the method of Section 3, we observe that f = likelihood prior takes the form

$$f(\alpha, \beta, \gamma, \sigma) = \alpha^{-1} \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha + \beta \gamma^{X_i})^2 \right\},$$

so that none of the four required complete conditional distributions are available in closed form. Hence four split-t envelopes will be needed.

For numerical illustration we use a data set from Ratkowsky (1983), displayed in Table 1, which tallies length (Y) and age (X) for 27 captured samples of the sirenian species dugong (more commonly known as the sea cow). To implement our method we first transform each of the variables to R^1 by letting $U_1 = \log \alpha$, $U_2 = \log \beta$, $U_3 = \text{logit}(\gamma) = \log(\gamma/(1-\gamma))$, and $U_4 = \log \sigma$. We then approximate the covariance matrix, Σ , of \underline{U} , using the quadratic regression approach mentioned above. We obtain four Cholesky matrices T from $\hat{\Sigma}$ by permuting the elements of $\hat{\Sigma}$ appropriately to make each of the U_i in turn the first element of \underline{U} . For this example we chose split-t distributions having $\nu = 5$ d.f.

[INSERT TABLE 1 ABOUT HERE]

Figures 1(a)–(c) show the marginal posterior density estimates for U_1 , U_2 and U_3 that result from the use of (1) on $m=500$ Gibbs iterates after completing $t=50$ iterations of the algorithm. We remark that density estimates on the original scales could be obtained via routine transformation as mentioned above (Gelfand and Smith, 1989). The posterior modes, .974, -.014 and 1.884 are comparable to the least squares estimates, .981, -.028 and 1.932 obtained by Ratkowsky (1983, p. 96).

4.2 Hierarchical Event Rate Model

To model arrivals or events occurring over known lengths of time we may use an exchangeable hierarchical model. For example, if Y_i is the number of occurrences over an exposure time of length t_i , $i=1, \dots, k$, we might assume that each Y_i is a realization from an independent Poisson process having constant rate λ_i , i.e., $Y_i \stackrel{\text{ind}}{\sim} P_0(\lambda_i t_i)$. We then assume that the λ_i are independent and identically distributed from some second stage distribution π . The conjugate choice for π would be a gamma distribution, so that the complete conditional distributions for the λ_i are updated gammas (see Gelfand and Smith, 1990). However, in order to allow for more dispersion and possible outliers in the rates, we might prefer a lognormal or logstudent-t prior for the λ_i , neither admitting closed form complete conditionals for the λ_i .

To develop the competing models more explicitly in the gamma case, we have at the

second stage $\lambda_i \stackrel{iid}{\sim} \text{gamma}(\alpha, \beta)$, $i=1, \dots, k$, where for convenience α is a known tuning constant. At the third stage of the hierarchy, we suppose $\beta \sim \text{IG}(c, d)$, where IG denotes the inverse (reciprocal) gamma distribution having mode $d/(c-1)$, and c and d are known constants. In the log-t case, letting $\epsilon_i = \log(\lambda_i)$ we have $\epsilon_i \stackrel{iid}{\sim} t_\omega(\theta, \sigma)$, where θ and σ are unknown location and scale parameters, respectively, and ω is specified (note that this parameter has nothing to do with ν , the degrees of freedom for our envelope split-t distribution). At the third stage of this model, we suppose $\theta \sim N(\mu, \tau^2)$ and $\sigma^2 \sim \text{IG}(a, b)$, θ and σ^2 independent, μ, τ^2, a and b known. Taking ω sufficiently large leads to the lognormal model for λ_i mentioned above.

Implementation of the Gibbs sampler is routine in the gamma case (see Gelfand and Smith 1990 for details). In the log-t case none of the $(k+2)$ necessary complete conditionals are standard distributions and hence, we apply the methods of Section 3. Since the likelihood factors into k pieces each involving only λ_i, θ and σ^2 , the remark at the end of Section 3 may be used to reduce the dimensionality of each of the first k component problems. However, to streamline the computer code we chose to ignore these savings, simply using the same $(k+2)$ -dimensional f function for each parameter under the parametrization $U_i = \epsilon_i = \log \lambda_i$, $i=1, \dots, k$, $U_{k+1} = \theta$, and $U_{k+2} = \log \sigma$. Here the covariance matrix is approximated using a derivative-free numerical Hessian.

The data in Table 2 are taken from Worledge, Stringham, and McClymont (1982), and record the number of failures of pumps over given lengths of time for several systems of a certain nuclear power plant. Gaver and O'Muircheartaigh (1987) also fit both the gamma and log-t models described above to this data, but employ an empirical Bayes approach, using the data to estimate all the parameters at the second stage of the model instead of placing third stage prior distributions on them. We make our analysis somewhat comparable in the case of the gamma by choosing $\alpha = \hat{\alpha} = 1.802$, the value of the method of moments estimator of α based on the marginal distribution of the data $m(Y | \alpha, \beta)$, and taking $c = 2.01$ and $d = 1.01$, so that β has prior mean 1 and prior standard deviation 10. In the log-t case, we specify the priors on θ by letting $\mu = -1, \tau^2 = 1$, and the prior on σ^2 by letting $a = 2.01$ and $b = 1.01$ (again, a rather vague prior with mean 1 and standard deviation 10). We use split-t distributions in our rejection algorithm, taking $\nu = 10$.

[INSERT TABLE 2 ABOUT HERE]

The estimated posterior distributions for ϵ_1 , ϵ_5 and ϵ_{10} under the gamma model, the log-t model with $\omega = 5$, and the log-t model with $\omega = 50$ (essentially a lognormal model) are shown in Figures 2(a) - (c) using (1) after $t = 30$ iterations with $m = 100$ replications. The results are similar to those of Gaver and O'Muircheartaigh (1987, p. 11). We see that, as expected, the gamma model generally produces posterior distributions that are more highly peaked and less dispersed. Also note that the gamma model seems to encourage more shrinkage to the grand mean of the observed rates. This is especially true for ϵ_5 , a rate corresponding to a system having a shorter history (smaller t_1).

4.3 Generalized Logistic Regression

As a final illustration we consider a class of models that find broad application in the social and biological sciences, especially in the context of dose-response studies. Suppose we have a Bernoulli response variable Z and a continuous predictor variable W . Typically one models the probability of a response at a given level of the predictor as

$$P(w) = \int_{-\infty}^y h(s) ds \quad (10)$$

where $y = (w - \mu)/\sigma$, μ and σ unknown. The most common assumption is to let h be the logistic distribution, which enables the closed form expression

$$P(w) = \exp y / (1 + \exp y), \quad (11)$$

i.e. the familiar logistic regression model. Prentice (1976) extended (11) introducing the class of models generated by taking

$$h(s) = \exp(sm_1)(1 + \exp(s))^{-(m_1+m_2)} / \beta(m_1, m_2), \quad m_1, m_2 > 0, \quad (12)$$

where $\beta(\cdot, \cdot)$ represents the beta function. Prentice remarks that with appropriate choice of m_1, m_2 other familiar models for binary response data emerge. More importantly he notes the potential improvement in fit afforded by the additional parameters. One convenient special case that enables an explicit form for $P(w)$ is to set $m_2 = 1$, obtaining

$$P(w) = [\exp y / (1 + \exp y)]^{m_1} \quad (13)$$

To effect a Bayesian analysis using model (13) we need to specify priors on μ , σ^2 and m_1 . Broad modeling possibilities arise by letting $m_1 \sim \text{Gamma}(a_0, b_0)$, $\mu \sim N(c_0, d_0)$, and $\sigma^2 \sim \text{IG}(e_0, f_0)$, m_1 , μ , and σ^2 a priori independent and a_0 , b_0 , c_0 , d_0 , e_0 and f_0 known. If we observe X_i responses out of n_i observations at predictor level w_i , $i=1, \dots, k$, f takes the form

$$f(\mu, \sigma, m_1) \propto$$

$$\left\{ \prod_{i=1}^k [P(w_i)]^{X_i} [1-P(w_i)]^{n_i-X_i} \right\} \exp \left\{ -\frac{1}{2} \left(\frac{\mu - c_0}{d_0} \right)^2 - \frac{m_1}{b_0} - \frac{f_0}{\sigma^2} \right\} \cdot (m_1^{a_0-1} / \sigma^{2(e_0+1)})$$

with $P(w)$ as given in (13). Again, the three complete conditional distributions will be sampled using the rejection method.

Our illustrative data set for this model is taken from Bliss (1935), and gives the proportion of adult flour beetles killed after five hours exposure to various levels of gaseous carbon disulphide. These data, displayed in Table 3, have been much-analyzed in the literature since their variability cannot be adequately explained by the standard logit model (11). We will compare the posterior distribution of μ ($=\text{LD}_{50}$ in the dose-response context) under the "full" model (13) and the "reduced" model (11).

[INSERT TABLE 3 ABOUT HERE]

For prior specification, we let $a_0 = .25$ and $b_0 = 4$, so that m_1 has prior mean 1 (the "reduced" value) and prior variance 4. We take rather vague priors on μ and σ^2 by letting $c_0 = 2$, $d_0 = 10$, $e_0 = 2.000004$, and $f_0 = .001$ (so that σ^2 has prior mean .001 and prior standard deviation .5). Using the obvious parametrization $U_1 = \mu$, $U_2 = \log \sigma$, and $U_3 = \log m_1$, empirical evidence suggests a multivariate split-t distribution with $\nu = 3$ to insure adequate domination of f in the tails. The covariance matrix is approximated using the quadratic regression approach.

Figure 3(a) shows the posterior distribution of LD_{50} under the full model (solid

line) and the reduced model (dashed line) arising from $t = 50$ iterations with $m = 500$ replications. The full model posterior mode of 1.815 is very close to the MLE value $\hat{\mu} = 1.818$ reported by Prentice (1976). The small overlap between the two posteriors in Figure 3(a) is consistent with the alleged lack of fit of the reduced model. Figure 3(b) shows similar problems in estimating $U_2 = \log(\sigma)$; its posterior distribution under the reduced model is also inappropriately centered and too highly concentrated. Still further evidence of the inadequacy of the reduced model is provided by the marginal posterior of $U_3 = \log(m_1)$ in Figure 3(c). Here we see the value assumed under the reduced model, $U_3 = 0$, is located in the extreme right hand tail of the estimated posterior.

5. Summary and Comments

To carry out calculations needed for Bayesian inference the Gibbs sampler is attractive in that, by utilizing complete conditional distributions, multivariate concerns become univariate ones. Moreover previous work (Gelfand and Smith, 1990; Gelfand et. al., 1989; Carlin et. al., 1989; Gelfand et. al., 1990) shows this approach to be a reasonably straightforward means for implementing fully Bayesian inference under conjugacy. This paper demonstrates that the Gibbs sampler can handle nonconjugate cases as well.

Nonetheless, problems which plague other techniques for Bayesian calculation (such techniques are discussed in Naylor and Smith, 1982, 1988; Smith et al, 1985, 1987; Tierney and Kadane, 1986; Geweke, 1989) will also cause difficulties for the Gibbs sampling approach. Such problems include disagreement between likelihood and prior, parametrization and flatness of the likelihood, strong posterior dependence amongst the parameters and of course high dimensionality. In such situations, successful use of the Gibbs sampler will require "tweaking". The severity of the above problems will dictate the extent of fine tuning required. Hopefully the conceptual simplicity of this iterative, univariate approach will simplify such matters.

We anticipate (and our examples support this) that the approach of Section 3 will be most effective in situations involving complicated likelihood but relatively low dimension or in higher dimensional situations as described at the end of Section 3 where for any parameter the number of other model parameters entering into its complete conditional distribution will be somewhat small.

Lastly we note that use of this rejection method by means of a distinct but convenient envelope for each iteration and each replication is described in Racine-Poon et. al., (1990). Extension of this approach certainly merits further investigation as does

comparison through challenging applications with the approach of this paper.

Acknowledgement

The authors acknowledge J. Berger for suggesting the use of Geweke's split distributions in the case of nonconjugate Gibbs sampling.

References

- Berger, J.O. (1985). Statistical decision theory and Bayesian analysis. Springer Verlag, New York.
- Berger, J.O. and Bernardo, J. (1989). Estimating a product of means: Bayesian analysis with reference priors. J. Amer. Statist. Assoc. 84, 200-207.
- Bernardo, J. (1979). Reference posterior distributions for Bayesian inference (with discussion). J. Royal Statist. Soc., Ser. B, 41, 113-147.
- Bliss, C.I. (1935). The calculation of the dosage - mortality curve. Annals of Applied Biology, 22, 134-167.
- Carlin, B., Gelfand, A. and Smith, A.F.M. (1990). Hierarchical Bayesian analysis of change point problems. Tech Rpt. 89-21, University of Connecticut, Dept. of Statistics.
- Devroye, L. (1986). Non-uniform random variate generation. Springer Verlag, New York.
- Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. Annals of Statistics, 7, 269-281.
- Gaver, D.P. and O'Muircheartaigh, I.G. (1987). Robust empirical Bayes analysis of event rates. Technometrics, 29, 1-15.
- Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M. (1989). Illustration of Bayesian inference in normal data models using Gibbs sampling. Tech. Rpt., University of Nottingham, Dept. of Mathematics.
- Gelfand, A.E. and Smith, A.F.M. (1989). Gibbs sampling for marginal posterior expectations. Tech. Rpt. 89-28, University of Connecticut, Dept. of Statistics.
- Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. J. Amer. Statist. Assoc. (to appear).
- Gelfand, A.E., Smith, A.F.M. and Lee, T-M. (1990). Bayesian analysis of constrained parameter and truncated data problems. Tech. Rpt. 90-01, University of Connecticut, Dept. of Statistics.

- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57, 1317-1339.
- Hills, S.E. (1989). The parametrization of statistical models. Unpublished Ph.D. thesis, University of Nottingham, Dept. of Mathematics.
- Kass, R.E. (1987). Computing observed information by finite differences. Communications in Statistics, Ser. B, 16, 587-599.
- Morris, C. (1983). Natural exponential families with quadratic variance functions: statistical theory. Annals of Statist., 11, 515-529.
- Naylor, J. and Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distributions. Applied Statistics, 31, 214-225.
- Naylor, J. and Smith, A.F.M. (1988). Econometric illustrations of novel numerical integration strategies for Bayesian inference. J. of Econometrics, 38, 103-126.
- Prentice, R.L. (1976). A generalization of the probit and logit model for dose response curves. Biometrics, 32, 761-768.
- Racine-Poon, A., Smith, A.F.M. and Gelfand, A.E. (1990). Bayesian analysis of population models using the Gibbs sampler. Tech. Rpt. 90-04, University of Connecticut, Dept. of Statistics.
- Ratkowsky, D. (1983). Nonlinear Regression Modeling. Marcel Dekker, New York.
- Ripley, B. (1988). Stochastic Simulation. J. Wiley & Sons, New York.
- Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C. and Dransfield, M. (1985). The implementation of the Bayesian paradigm. Communications in Statistics, Theory and Methods, 14, 1079-1102.
- Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C. (1987). Progress with numerical and graphical methods for Bayesian statistics. The Statistician, 36, 75-82.
- Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. J. Amer. Statist. Assoc., 81, 82-86.
- Worledge, D.H., Stringham, R.S. and McClymont, A.S. (1982). PWR power plant reliability data. Interim Report NP-2592, Electric Power Research Institute, Palo Alto.
- Zeger, S.L. and Karim, M.R. (1989). Generalized linear models with random effects; A Gibbs sampling approach. Tech. Rpt. P691, John Hopkins University, Dept. of Biostatistics.

Table 1. Length (Y) Versus Age (X) for the Sirenian Species Dugong

| | | | | | | | | | |
|---|------|------|------|------|------|------|------|------|------|
| X | 1 | 1.5 | 1.5 | 1.5 | 2.5 | 4.0 | 5.0 | 5.0 | 7.0 |
| Y | 1.80 | 1.85 | 1.87 | 1.77 | 2.02 | 2.27 | 2.15 | 2.26 | 2.35 |
| X | 8.0 | 8.5 | 9.0 | 9.5 | 9.5 | 10.0 | 12.0 | 12.0 | 13.0 |
| Y | 2.47 | 2.19 | 2.26 | 2.40 | 2.39 | 2.41 | 2.50 | 2.32 | 2.43 |
| X | 13.0 | 14.5 | 15.5 | 15.5 | 16.5 | 17.0 | 22.5 | 29.0 | 31.5 |
| Y | 2.47 | 2.56 | 2.65 | 2.47 | 2.64 | 2.56 | 2.70 | 2.72 | 2.57 |

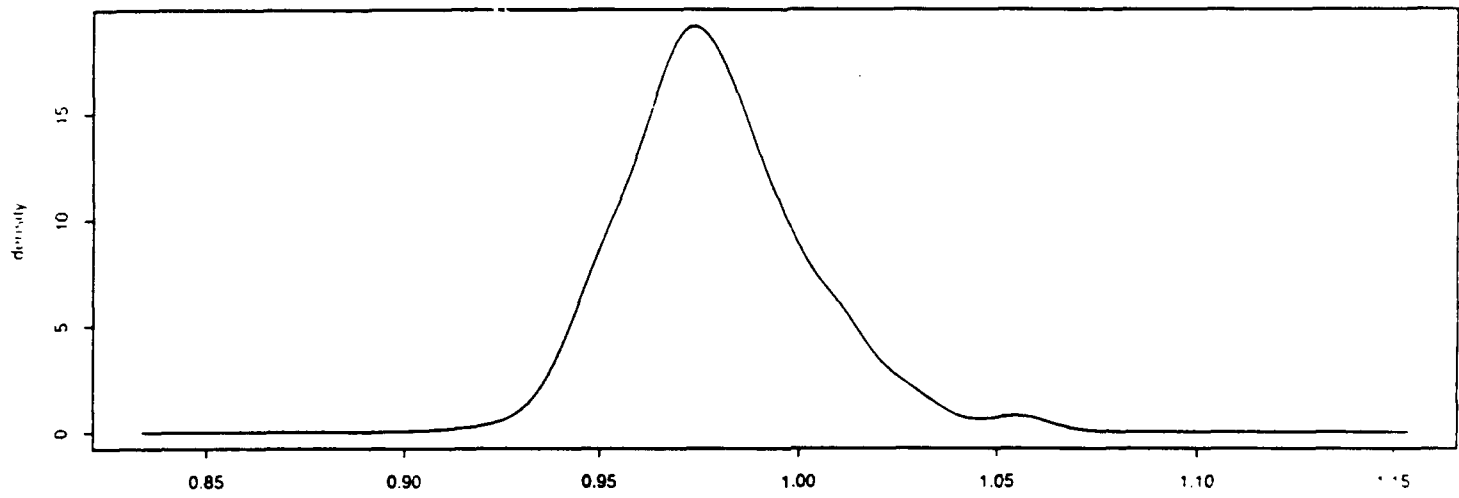
Table 2. Pump Failures (t_j in thousands of hours)

| System j | X_j | t_j |
|------------|-------|---------|
| 1 | 5 | 94.320 |
| 2 | 1 | 15.720 |
| 3 | 5 | 62.880 |
| 4 | 14 | 125.760 |
| 5 | 3 | 5.240 |
| 6 | 19 | 31.440 |
| 7 | 1 | 1.048 |
| 8 | 1 | 1.048 |
| 9 | 4 | 2.096 |
| 10 | 22 | 10.480 |

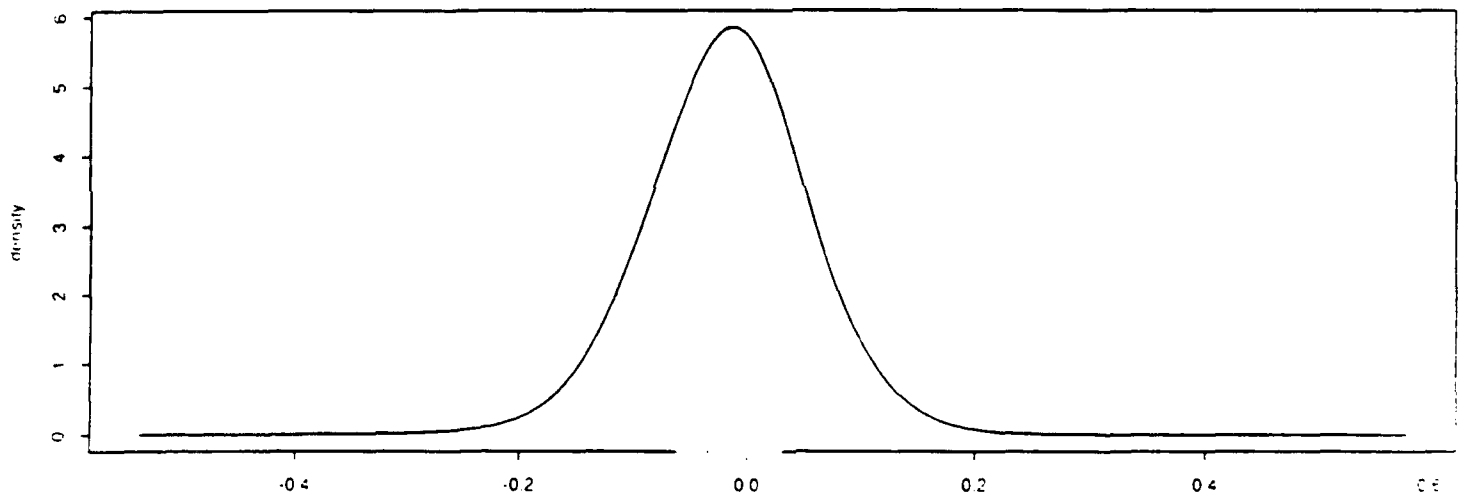
Table 3. Observed Flour Beetle Mortality Data

| Dosage CS_2 | No. of Beetles | |
|------------------|----------------|---------|
| | Killed | Exposed |
| 1.6907 | 6 | 59 |
| 1.7242 | 13 | 60 |
| 1.7552 | 18 | 62 |
| 1.7842 | 28 | 56 |
| 1.8113 | 52 | 63 |
| 1.8369 | 53 | 59 |
| 1.8610 | 61 | 62 |
| 1.8839 | 60 | 60 |

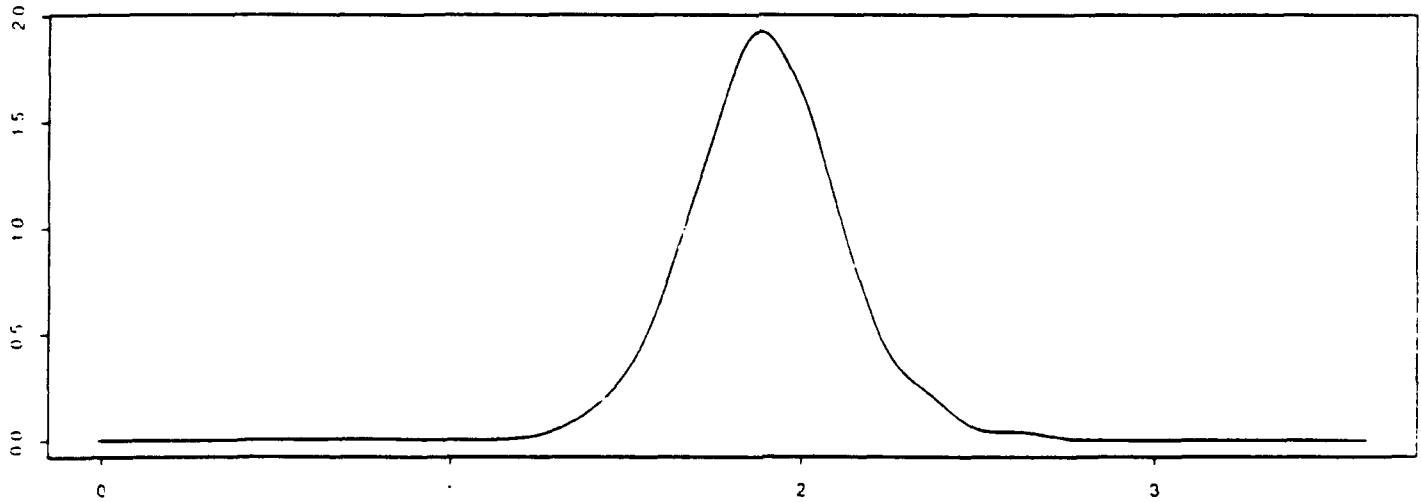
Figure 1. Estimated posteriors, dugong data



a) Marginal posterior for $\log(\alpha)$, $m = 500$; mode = 0.974

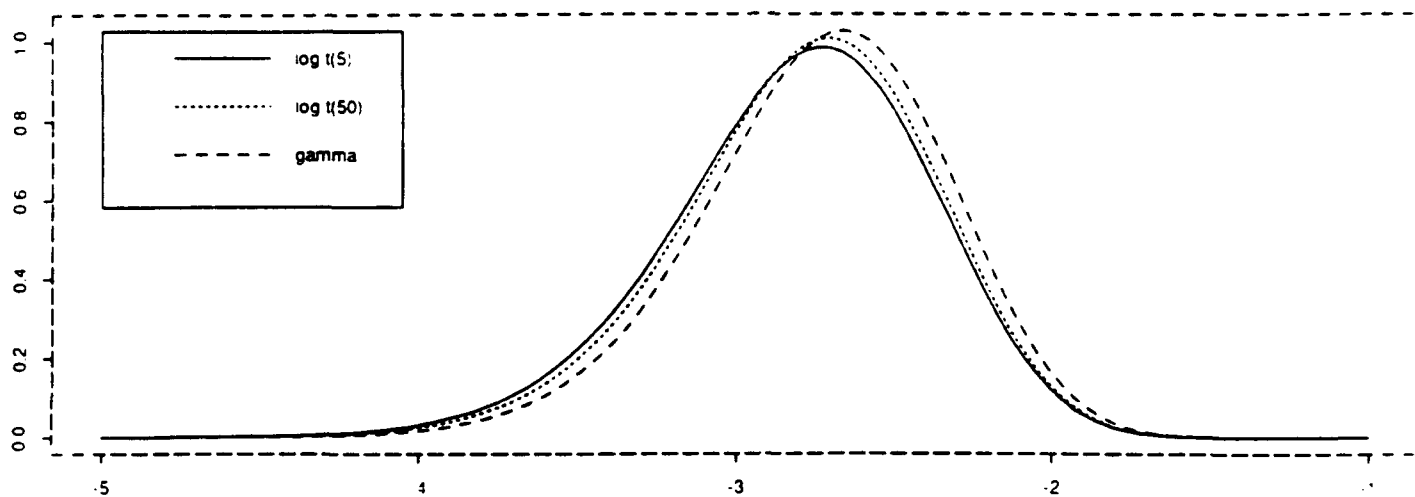


b) Marginal posterior for $\log(\beta)$, $m = 500$; mode = -0.014

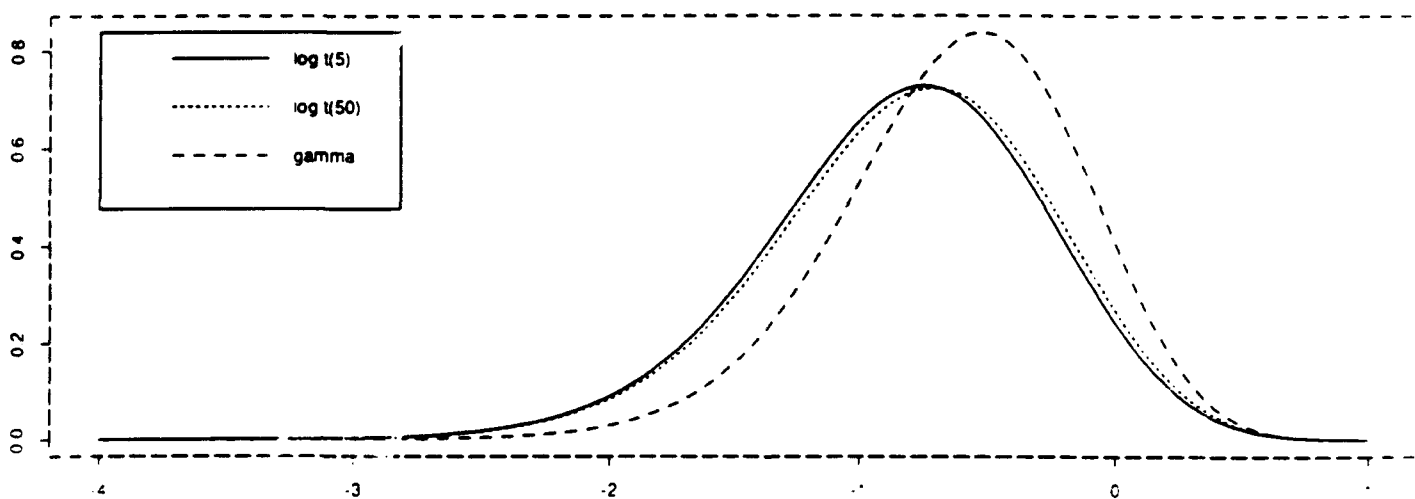


c) Marginal posterior for $\text{logit}(\gamma)$, $m = 500$; mode = 1.884

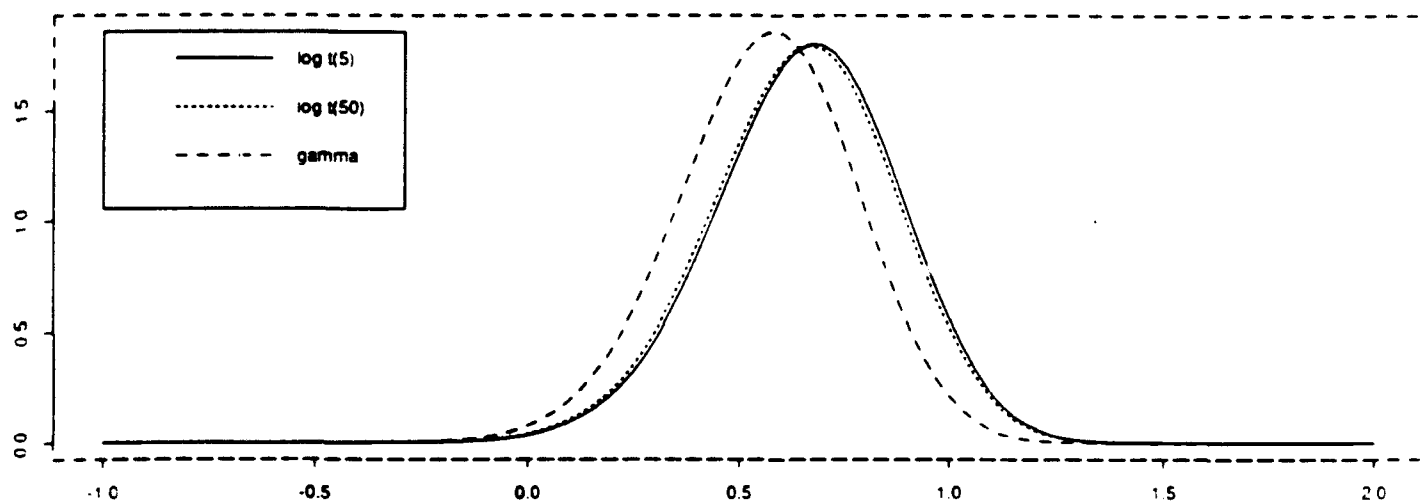
Figure 2. Estimated posteriors, pump failure data



a) Marginal posterior for $\log(\lambda_1)$, $m = 500$

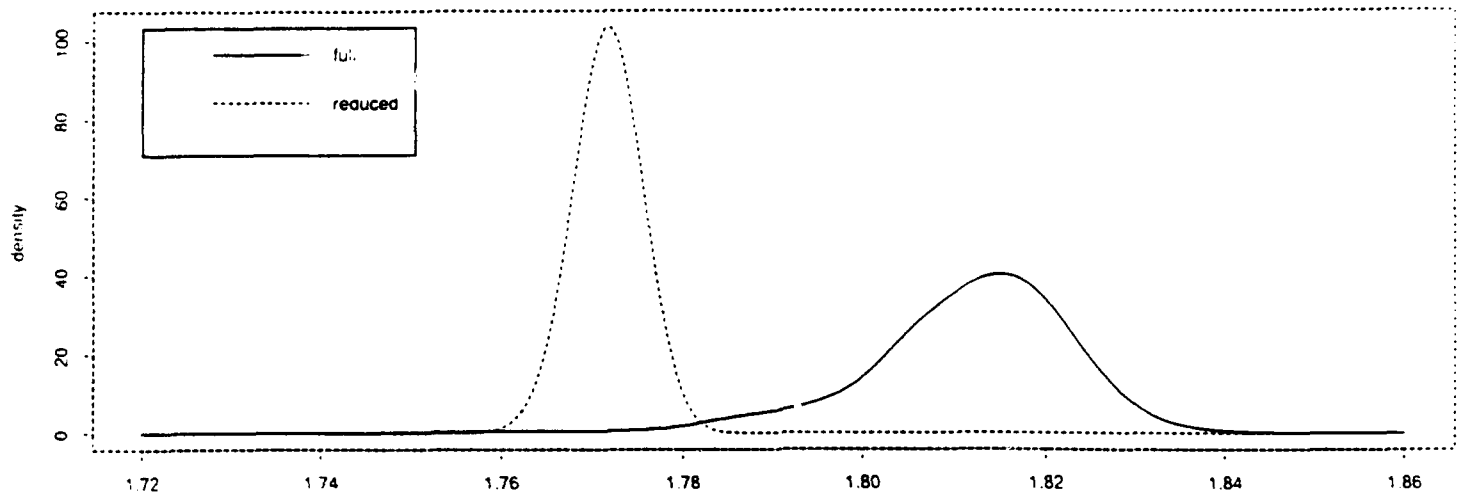


b) Marginal posterior for $\log(\lambda_5)$, $m = 500$

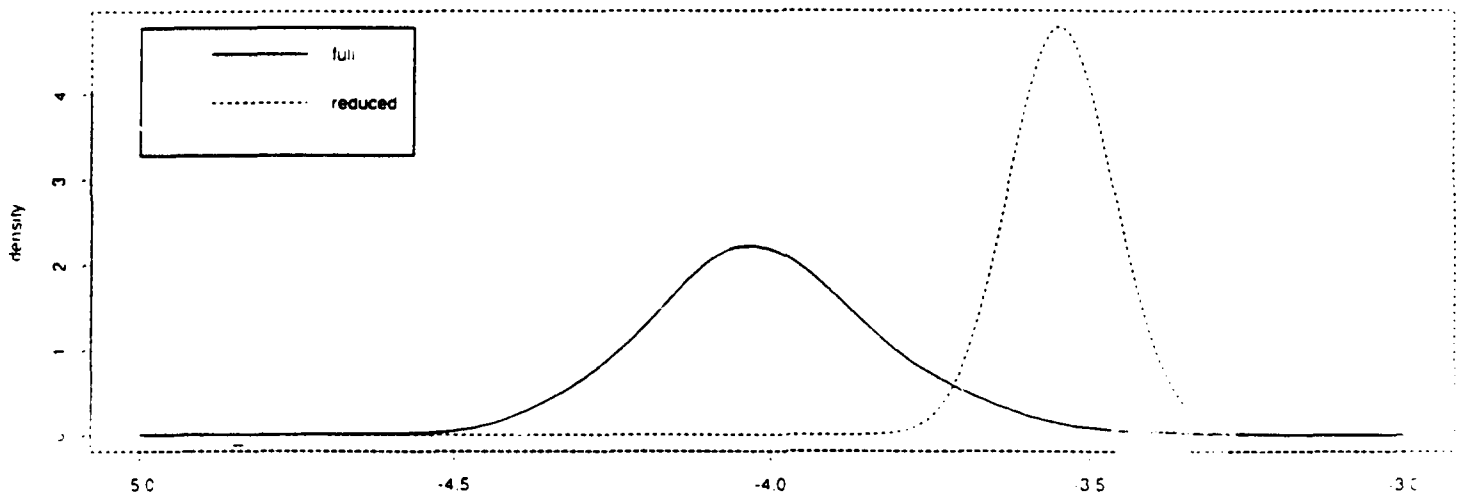


c) Marginal posterior for $\log(\lambda_{10})$, $m = 500$

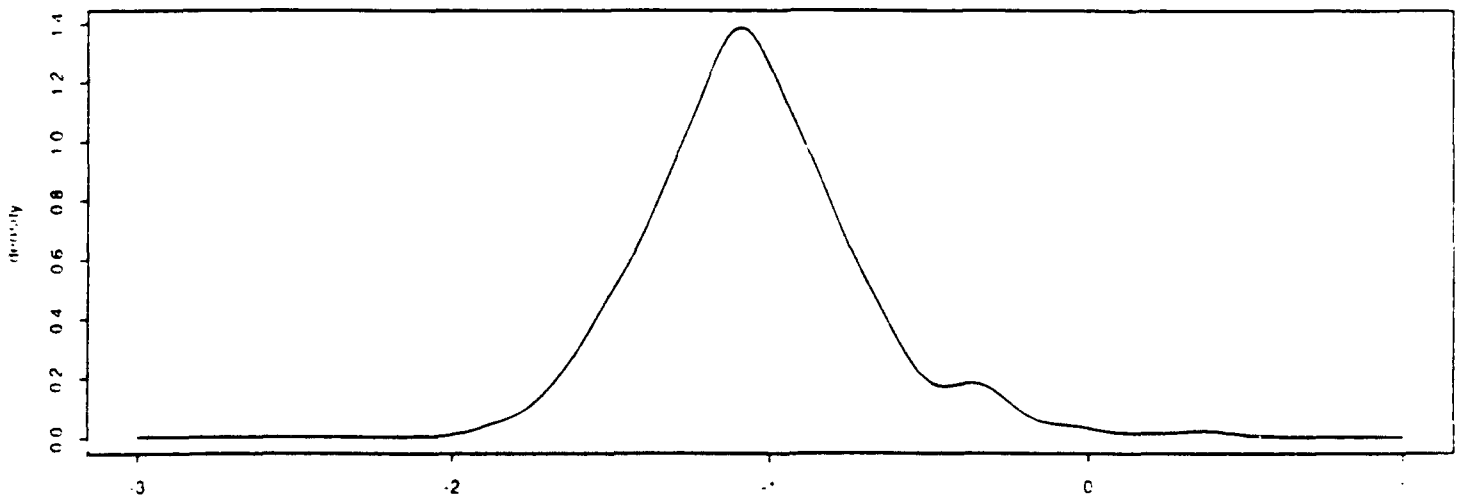
Figure 3. Estimated posteriors, beetle data



a) Marginal posterior for LD50, $m = 500$



b) Marginal posterior for $\log(\sigma)$, $m = 500$



c) Marginal posterior for $\log(m1)$, $m = 500$, mode = -1.05

ABSTRACT

The Gibbs sampler has been proposed as a general method for Bayesian calculation in Gelfand and Smith (1990). However experience to date is almost exclusively in applications assuming conjugacy where implementation is reasonably straightforward. This paper describes a tailored rejection method approach for implementation of the Gibbs sampler when nonconjugate structure is present. Several challenging applications are presented for illustration.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|--|-----------------------|--|
| 1. REPORT NUMBER 457 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) AN ITERATIVE MONTE CARLO METHOD FOR NONCONJUGATE BAYESIAN ANALYSIS | | 5. TYPE OF REPORT & PERIOD COVERED Technical |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) Bradley P. Carlin and Alan E. Gelfand | | 8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code: 111 | | 12. REPORT DATE September 7, 1992 |
| | | 13. NUMBER OF PAGES |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number) <i>Keywords: Bayesian inference. Gibbs sampler. hierarchical models. logistic regression. nonlinear models, rejection method</i> | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) See Reverse Side | | |